Transactions of the American Fisheries Society 119:41–54, 1990

Use of a Computer Simulation Model to Determine the Behavior of a New Survey Estimator of Recreational Angling

CYNTHIA M. JONES[1]

U.S. Fish and Wildlife Service, Cooperative Fish and Wildlife Unit
Cornell University, Ithaca, New York 14853, USA

DOUGLAS ROBSON[2]

Biometrics Unit, Cornell University

DAVID OTIS[3] AND STEVEN GLOSS[4]

U.S. Fish and Wildlife Service, Cooperative Fish and Wildlife Unit, Cornell University

Abstract.—We have recently developed a new estimator of recreational fishing effort. Here we used a simulation model to determine and demonstrate the statistical behavior of this new estimator. The estimator is used with access-intercept surveys and was designed to give accurate, efficient estimates of fishing effort over a geographically large and diverse fishery. Because this is a new estimator, little is known about its behavior. Specifically, the form of the estimator's sampling distribution, the variance components (within-day versus between-day), and its t-distribution were unknown and could not be determined analytically. Hence, to assist people who will want to know the statistical properties of this estimator and to characterize it more completely, we studied its behavior numerically by use of a simulation model based on real-world data. Analysis of the simulation results showed the sampling distribution of the estimator to be non-normal when limited to a single survey route; it was more closely approximated by a gamma distribution. The estimator approached normality when used to estimate effort from multiple-route (large-scale) fisheries with greater fishing effort. Within-day variance (influenced by starting position along the route and direction of travel) was larger than the between-day (day-to-day) variability. Because the sampling distribution of the estimator was non-normal, the t-distributions were generated empirically to determine the direction and degree of misspecification when the usual Student t-distribution was used. Use of the Student t resulted in slightly skewed α values with too large a probability of inclusion in the lower tail and too small a probability in the upper tail.

In 1984, the New York Department of Environmental Conservation (NYDEC) undertook an extensive contact angler survey of the fisheries on New York's Great Lakes. Traditional survey designs were inadequate because there were too many access sites. The survey, which covered the New York portions of the lakes and their associated tributaries, spanned a full year of fishing over all major types of recreational angling (tributary, boat, ice, etc.). The constraints on the survey design were those of budget and personnel. There were too few survey agents available to cover all of the access sites when a standard access method was used. The survey route design needed to cover the maximum area with the most efficient use of personnel and was developed to be both time- and labor-efficient and to yield unbiased, precise estimates of fishing effort on a lakewide basis. The NYDEC needed these estimates of fishing effort to evaluate their restoration program for lake trout *Salvelinus namaycush* and to assess the impact o salmonid stocking.

To meet the needs of this survey, a new est mator of recreational fishing effort (and catch) wa developed (Robson and Jones 1989). The ne procedure is analogous to a bus route. Instead visiting just one or two access sites a day (t traditional approach), each survey agent make complete circuit of all access sites each sampl day. The agent has a precise schedule to foll each day and arrives and departs from each

[1] Present address: Applied Marine Research Laboratory, Old Dominion University, Norfolk, Virginia 23508, USA.

[2] Present address: 150 McClaren Street, Penthouse 6, Ottawa, Ontario K2P 0L2, Canada.

[3] Present address: U.S. Fish and Wildlife Service, Denver Wildlife Research Center, Denver, Colorado 80225, USA.

[4] Present address: Wyoming Water Research Center, University of Wyoming, Post Office Box 3067, University Station, Laramie, Wyoming 82071, USA.

on a prespecified timetable. Because the starting point along the route is chosen randomly each day, each site is visited randomly throughout the day over the survey period. If, for example, there were 12 parking lots to visit, it would be unreasonable to spend a full day at one of so many access sites. Each site would be sampled infrequently within a month. With our new estimator, we use one crew of agents to cover all 12 sites within a day. With a standard access design, we would have needed one crew to cover one or two sites per day, or 6–12 crews per day. Additionally, the new estimator can be used to estimate fishing effort in situations in which the parties' cars are seen, but the likelihood of obtaining an interview is small. The length of time that an angling party's car is at an access site while the survey agent is present gives an unbiased estimate of fishing effort.

Part of the responsibility in bringing out a new estimator for use by the fishery professional is to provide information about the estimator so that variances and confidence intervals can be calculated. Because our estimator used a totally new design, little was known about its behavior before our numerical modeling. It was already known that the estimator was unbiased. However, the properties and shape of the sampling distribution were unknown and could not be determined analytically.

To understand the application of the numeric models, one needs to be familiar with characteristics of the estimation procedures. A complete explanation of the survey methodology and derivation of estimators used in the survey are given in Robson and Jones (1989). The angler survey obtained an estimate of fishing effort on a lakewide basis. This estimated total daily fishing effort for a given survey route, ETPH (estimated total party hours), is

$$\text{ETPH} = T \sum_i \left(\frac{1}{w_i}\right) \sum_j X_{ij};$$

T = total time for a survey agent to travel the route;

X_{ij} = amount of time angler party's car (j) is present during the survey agent's wait at site i; and

w_i = the survey agent's waiting time at the ith access site. Waiting time may be of equal duration for all sites within a tributary route.

Note that in the Robson and Jones (1989) paper, ETPH was given the notation T_{ph}. During a time stratum (the strata were weekdays, weekend days, and holidays), the survey was conducted on a random sample of n days. For example, 8 weekdays were selected at random from among the 20 weekdays available per month. The variable ETPH was calculated separately for each day, and the mean ETPH was calculated for each month. The unbiased estimator for this mean ETPH ($\overline{\text{ETPH}}$) during that month is

$$\overline{\text{ETPH}} = \frac{1}{n} \sum \text{ETPH}.$$

The sampling error variance estimator is a sum of the within-day and between-day variance components, arising from the relationship

$$\overline{\text{ETPH}} - \overline{\text{ATPH}} = \frac{1}{n} \sum_{k=1}^{n} (\text{ETPH}_k - \text{ATPH}_k) + \frac{1}{n} \sum_{k=1}^{n} (\text{ATPH}_k - \overline{\text{ATPH}});$$

ATPH (actual total party hours) is the actual number of party hours that are fished on a route for a given day (k), and $\overline{\text{ATPH}}$ is the mean of the actual party hours fished on a route calculated for each month. (This is the same as equation 14 in Robson and Jones [1989].) The within-day error component of daily ETPH_k, expressed as $(1/n)\sum(\text{ETPH} - \text{ATPH})$ in the above equation, reflects variation due to the position of the agent's random start along the route and its effect on encountering parties. The between-day error component, expressed as $(1/n)\sum(\text{ATPH} - \overline{\text{ATPH}})$ in the above equation, is due to day-to-day changes in the angler population. However, for the field survey, only total variance was estimated, and no estimate was made for within-day variance because only one crew of survey agents was used along a route each sampling day. To estimate within-day variance would have necessitated several crews conducting two or more randomized surveys along a given route on the same day, a costly procedure in this case or in the conventional access site survey. Hence, variance components (i.e., within-day versus between-day) could not be separated. For the 1984 field study, we chose a conservative variance estimator for building confidence intervals about the estimates of effort. This conservative estimator is given in equation (18) in Robson and Jones (1989):

$$v(\text{ETPH}) = [1/n\,(n - 1)]\sum(\text{ETPH} - \overline{\text{ETPH}})^2 = (1/n)s^2_{\text{ETPH}};$$

v = total variance of $\overline{\text{ETPH}}$, and s^2_{ETPH} = sample estimate variance. Use of a moderately conservative variance estimator results in variance estimates that are, on average, larger than the true variance and that yield larger confidence intervals.

Our simulation model was developed to assess the behavior of this new survey estimator. Although the estimator itself was shown to be unbiased (Robson and Jones 1989), the shape of the sampling distribution of the estimator and hence the variance and t-values were unknown. This ability to estimate variance and calculate the t-values was essential for developing confidence intervals. The simulation model constructed was based on information obtained in field interviews. This was done with a process similar to the "bootstrapping" procedure (Efron 1979a), in which the population of interest is built or tested by emulating real-life procedures and basing the model parameters on real-world data (Efron 1979b). This procedure permits the construction of confidence intervals that otherwise would not be known (Diaconis and Efron 1983).

The numeric model for this study was built at considerably less cost than that of performing additional field surveys. The examples that we will use represent fishing along a stream. However, the behavior of the estimator is consistent over all fisheries, and the results are applicable to other types of fisheries (C. M. Jones, W. Check, A. Ehtisham, and P. Geer, Old Dominion University and K. H. Pollock, University of North Carolina, unpublished data). The model was used to investigate the sampling distribution of the estimator and the shape of the t-distribution, information that cannot be obtained with any other technique.

Here we show the statistical properties of this new estimator in a thorough manner, including the shape of the sampling distribution of the estimator compared with that of the actual population under weekday and weekend-day sampling scenarios, and when estimating single and combined tributary routes. Additionally, we illustrate a unique and useful application of simulation modeling for developing new recreational angling survey designs.

Construction of the Model

The Computer Model Components

The computer model had three modules: (1) the angler population module, (2) the estimation module, and (3) the output analysis module (Figure 1).

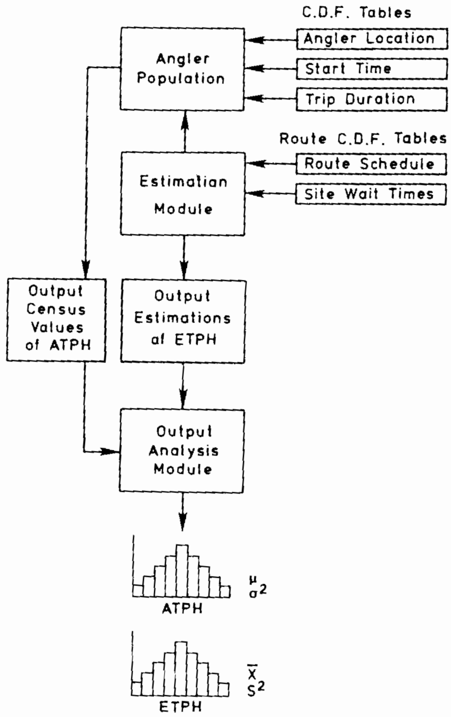

FIGURE 1.—Components of the computer simulation model. The model will allow input for route designs that are user specified and uses the characteristics (e.g., starting times) specific to the user's fishery through cumulative distribution function (C.D.F.) tables. ATPH = actual total party hours of fishing; ETPH = the estimated total party hours; μ and σ^2 are the mean and variance of the population; $\bar{x}$ and s^2 are the mean and variance of the sample.

The angler population module stochastically assigned a user-chosen number of fishing parties to access sites and assigned starting time and duration for each party's trip on the fishery. The input data to this module were information obtained from completed trip interviews from the 1984 New York Great Lakes angler survey, but could be comparable data from any survey. This information was input from user-designated arrays or files. The output data were the actual number of party hours (ATPH) assigned for each iteration or "day" of fishing simulated by the model.

The estimation module simulated the survey agent's travel around a route. The input data to the estimation module were the characteristics of the route, e.g., the number of access sites and the route schedule (length of travel between sites and waiting times at each site). This information was input from user-specified files at the beginning of the simulation run. The output of the estimati module was the estimated total party hours (ETPH) for each iteration. A direct comparison with each iteration's output of ATPH from the angler population module can be made.

The output analysis module analyzed the output from the angler population module and estimation module by use of PC Minitab and other custom programs. The output files from the angler population module and the estimation module are in one-to-one correspondence. Therefore, the difference between the estimated value of fishing effort and the actual value can be obtained. Custom-designed programs were used to subsample the simulation output to represent sampling choices taken from weekend-day and weekday strata. Programs were also designed to calculate the empirical t-value.

Creation of the Angler Population Module

Design of the survey route.—This new estimator was designed to estimate effort for an individual survey route. The survey route design is analogous to a bus route, which is built on a precise time schedule (Figure 2). Each day, the starting point along the route was chosen randomly. Agents proceeded around each route, and arrived and departed on a precise time schedule. While the agent waited at an access site, the amount of time that an angling party's car was in view was recorded as a measure of fishing effort. Because fishing effort was heavier on weekends, the survey was stratified into weekday and weekend-day sampling periods. The daily sampling schedule was chosen randomly to start at either the beginning of the fishing day (an early start) or to end at the end of the fishing day (a late start).

Choice of emulated routes.—We chose the Sterling River and Upper Salmon River survey routes from the New York Great Lakes survey for emulation. These routes were representative of low and high fishing pressure, respectively. For the Sterling River route simulation, the numbers of fishing parties were set at 10, 20, 50, 100, and 200. This produced a reasonable range of fishing effort for a low-effort route. The numbers of parties chosen for the Upper Salmon River route were 10, 20, 50, 100, 200, 500, and 1,000. This produced a reasonable range of fishing effort for a high-effort route. One thousand iterations of the simulation model were used in analysis.

Building cumulative distribution function tables.—The simulation model generated random values for party location, starting time, and trip duration from user-specified distributions. Party location and starting time were chosen independently. Actual completed trip interviews provided the basic data upon which continuous, piecewise linear empirical distribution functions were built (Law and Kelton 1982). It was known that the estimators for these input variables were unbiased (Robson and Jones 1989). The cumulative empirical distribution functions were used in conjunction with a uniform random number generator to generate the random variates used in the angler population module (Morgan 1984).

Party location.—We obtained the relative distribution of parties at access sites along the survey route from the field survey and used it to apportion parties to access sites within the simulation model. For example, if site 1 had 30% of the field interviews from the actual field survey, then over many iterations of the model, 30% of the parties would have been assigned at this site for the simulated populations. This resulted in n_i angling parties at the ith site. For a given iteration, this would not necessarily be true.

Starting time of the party excursion.—We used information obtained from completed trip interviews to construct a cumulative distribution function of angler starting times (Figure 3). The two cumulative distribution functions of starting times from the weekday and weekend-day strata were virtually identical. Based on this empirical cumulative distribution function, each of the n_i parties at the ith site was stochastically assigned a starting time.

Trip duration.—Trip duration was dependent on angler starting time. Anglers starting early in the day fished longer than those starting later (Figure 4). For stream fishing, the relationship between starting time and mean trip duration was linear. After a starting time was simulated, we determined trip duration by choosing a uniform random number between 0 and 1 and multiplying it times the number of minutes remaining in the fishing day. Time of completion of the angler excursion was obtained by adding the starting time to the trip duration.

Summary of the angler population module.—We summed the stochastically assigned trip duration for all parties during each iteration of the model. This provided the "ground truth" to which the estimation of fishing effort could be compared. The attributes of party location, and starting and ending times were used in the estimation module query of the angler population module to determine whether the survey agent and angling party overlapped in time and space.

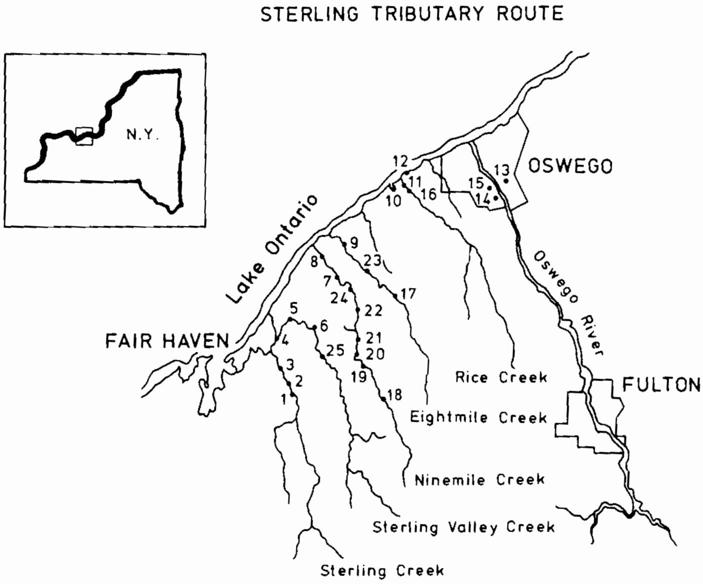

FIGURE 2.—Illustration of the Sterling River, New York, tributary route. The route was followed at a randomly chosen starting point, either by increasing (clockwise) or decreasing (counterclockwise) site number.

The Estimation Module

Building the daily route.—Input to the estimation module was the route schedule (i.e., cumulative minutes traveling to each site plus the waiting times). The starting point along the route and the direction of travel were chosen at random for each iteration. Additionally, a binary random number (essentially a heads or tails choice) was used to determine whether the simulated route covered the beginning of the fishing day (early start) or the end of the fishing day (late start) for each iteration. Once the starting point and starting time were determined, the daily schedule was built for the route. The time of simulated agent arrival and departure for each site was stored for each iteration.

Data collection.—At each site, the simulated agent arrival and departure times at the site were compared to the start and finish times for the sim-

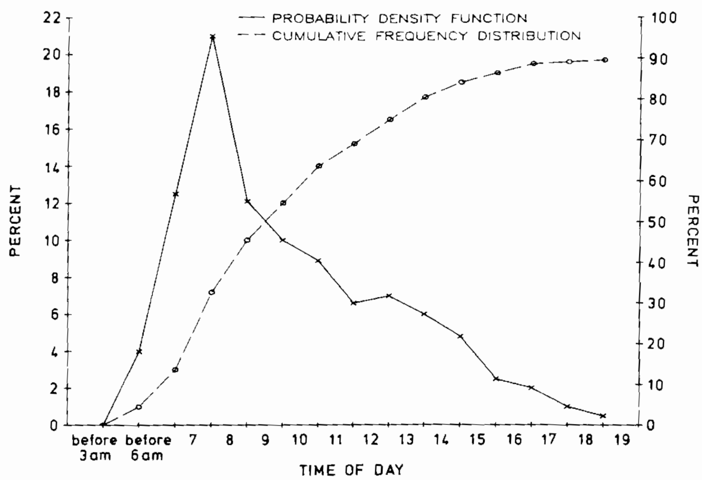

FIGURE 3.—Probability density function (left ordinate axis) and cumulative distribution function (right ordinate axis) for tributary angler starting times. The starting times for anglers fishing on weekend days and weekdays were pooled because they were not significantly different. Time is in hours (7 = 0700 hours).

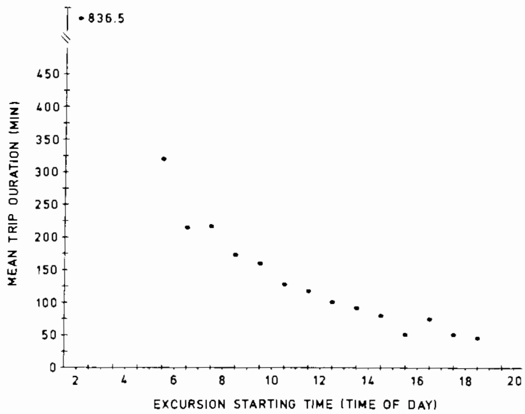

FIGURE 4.—Duration of tributary fishing excursions in relation to the time that the excursion started. When overnight fishing was excluded, trip duration was closely modeled with a linear function. Time is expressed in hours (2 = 0200 hours).

ulated angler party. For those parties at the site concurrently with the agent, the amount of time that the party was present during the waiting period was added to the cumulative estimated fishing effort total (ETPH) for the route for each iteration or "day." By the end of each iteration ("day"), the estimation module produced an estimate of total fishing effort to compare with the output of the angler population module. Complications arise when the survey sampling time is greater than half but less than the total length of the fishing day. Under these circumstances, a middle period of the sampling day resulted that was sampled each survey day, whereas the early or late parts of the day were sampled only every other sampling day (Robson and Jones 1989). This is compensated for by weighting the early and late parts of each sampling day twice as heavily as the middle period.

Output Analysis Module and Statistical Evaluation

Evaluation of actual and estimated fishing effort.—The sum of party hours from all simulated angler excursions (ATPH) produced for each iteration ("day") of the angler population module and the sum of estimated party hours (ETPH) produced for each iteration ("day") of the estimation module were output at the end of all iterations. These values were used to calculate the difference (DTPH) between each matched value of actual (ATPH) and estimated (ETPH) party hours. We produced histograms for ATPH, ETPH, and DTPH, along with their respective means and standard errors. Histograms of ETPH and DTPH showed the sampling distributions of the estimator and of the differences between the actual and estimated values. This was done for single-tributary cases and for combined Sterling River and Upper Salmon river tributary routes. The values for the two single-tributary routes were combined to represent effort as calculated in field studies. The new estimator was designed to estimate effort on a multiple-route basis. To represent weekend days, the combined data from the Sterling River 200-party and Salmon River 1,000-party iterations were used to represent the high fishing effort seen on weekends. Because both weekend days were sampled each week, a data base was created that was broken into 1,000 8-d segments, each representing a month of weekend sampling. The values (means) of each segment were plotted as histograms and used to calculate t-values. The calculations for weekdays differed slightly from weekend days because only 8 days out of a possible 20 weekdays/month were sampled in the field. The simulation model reflected this difference. We created a data base of 1,000 sets of 20 iterations. The values of ETPH and ATPH were chosen from 8 iterations randomly from each set of 20. The mean values of these eight-iteration segments were plotted, and the values used to calculate the t-statistics.

Analysis of combined tributary routes.—We performed a further simulation to demonstrate the behavior of the estimator during the performance of its intended design (i.e., that of calculating estimates of fishing effort over an entire fishery made up of two or more routes). For this purpose, we chose a combination of the Sterling River tributary route with 10 parties and the Upper Salmon River tributary route with 100 parties for illustration. The simulation procedure to parallel weekend-day sampling used 1,000 iterations of 8 d, combining a day each from the Sterling and Upper Salmon River routes chosen randomly without replacement from 1,000 potential sampling days. The simulation procedure that paralleled weekday sampling used 1,000 sets of 8 iterations chosen at random without replacement from sets of 20 iterations. This reflected field situations when 2 d were sampled out of 5 weekdays each week for a selection of 8 out of 20 potential sampling days for each month.

Evaluation of confidence interval coverage.—Values of t were calculated as

$$t_i = (\mathrm{ETPH}_i - \mu_i)/\mathrm{SE(ETPH)};$$

μ_i is the monthly mean value of either the weekend-day or weekday ATPH, and SE is the standard error of the estimate, depending on the appropriate application. Values of t were computed for the combined Sterling River and Upper Salmon River tributary routes, both for weekend-day and weekday simulations. We examined the distribution of computed t-values to determine the functional form to be used in confidence interval estimation.

Results

Behavior of the Estimator for Singular Tributary Routes

Sterling River tributary route with 10 parties.— The simulations for a single route showed that the sampling distribution of our new estimator was non-normal. Results from the simulation of the Sterling River route with 10 parties are used as an example in the text and figures. Results from the simulations of the Sterling River route with 20, 50, 100, and 200 parties are presented in Table 1. Figure 5 shows the results of a single "day's" generation of 10 angling parties distributed among 25 access sites on the Sterling River route with starting times and fishing duration selected randomly from the empirical distributions (cumulative distribution functions). This collection of parties and values of fishing effort, totaling to ATPH = 40.5, remained fixed during the first set of simulations (Figure 5A). Meanwhile, we used 1,000 different randomly chosen starting points along the route and subsequent simulated survey runs to generate 1,000 different estimates of fishing effort, ETPH (Figure 5B). The mean of the 1,000 estimates of ETPH was 40.5. The estimates ranged from 0 to 90 total party hours. Figure 5C illustrates the one-day difference between ETPH and ATPH, or DTPH, which is simply the value for each individual ETPH minus the true value of ATPH, 40.5. This figure demonstrates the magnitude of the within-day variability.

We generated the histogram of ATPH for 1,000 "days" (Figure 6A) from 1,000 collections of 10 angler parties independently simulated exactly as for the single day shown in Figure 5A. The simulation of ATPH resulted in a mean of 38.3 total party hours. One randomly started survey run was simulated on each of the same 1,000 "days" to produced a histogram of 1,000 ETPH values (Figure 6B). The mean of these 1,000 values of ETPH, 38.4, is an unbiased estimate of the mean of 1,000 values of ATPH, 38.3. Estimates of ETPH have

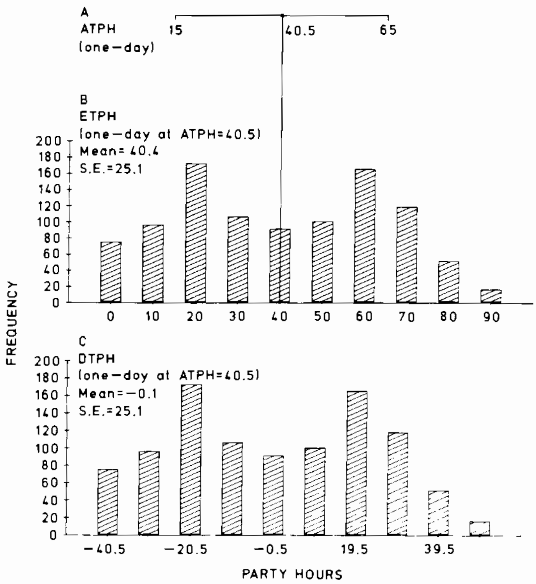

FIGURE 5.—Simulated behavior of the estimator of fishing effort of 10 parties for "one day" on the Sterling River tributary. (**A**) Values for angling hours which were stochastically produced for one "day" or iteration of the model resulted in a mean actual total party hour (ATPH) value of 40.5. (**B**) One thousand different starting locations and directions produced 1,000 estimates of total party hours (ETPH). (**C**) Difference (DTPH) between the estimated and actual party hours shows the magnitude of within-day variability.

a broader range compared with ATPH because the variance of ETPH combines the within-day estimation error of Figure 5C and the day-to-day variability in 1,000 estimates of ATPH (Figure 6A). Also note the skewness in the distribution of ETPH (Figure 6B) similar to that seen in Figure 5B. The skewness of this distribution is mainly driven by the skewness of the within-day component. The histogram of DTPH (Figure 6C) illustrates the within-day estimation error; each day's ATPH is subtracted from the same day's ETPH. The histogram of these 1,000 estimation errors (Figure 6C) represents 1,000 samples of size 1 from each of 1,000 distributions of DTPH of the type illustrated for the single distribution of Figure 5C. The implication from the greater range for DTPH in Figure 6C compared with the range in Figure 5C is that some of the distributions of the type shown in Figure 5C were broader than this specific example. The variance (σ^2_{DTPH} = 750.76) of the distribution (Figure 6C) estimates the average (mean σ^2_{DTPH}) of all 1,000 of the variances illustrated by a single variance (σ^2_{DTPH} = 630.01 for ATPH = 40.5) shown in Figure 5C.

TABLE 1.—Means and variance components from simulations of the Robson and Jones (1989) estimator for the Sterling River and Upper Salmon River tributaries in New York State. CV = coefficient of variation (100 × SE/mean); ATPH = actual total party hours of fishing; ETPH = estimated total party hours of fishing; SE(DTPH) = average within-day variability computed as the square root of the (variance of ETPH minus the variance of ATPH).

Number of parties	(1) ATPH Mean	(2) ATPH SE	(3) ATPH CV	(4) ETPH Mean	(5) ETPH SE	(6) ETPH CV	(7) DPTH SE	(8) Percent of total variation[a]
Sterling River								
10	38.3	8.5	22.2	38.4	28.6	74.5	27.3	91.1
20	77.4	12.4	16.0	76.6	49.5	64.6	47.9	93.6
50	192.7	19.2	10.0	197.9	114.1	57.7	112.5	97.2
100	384.5	27.3	7.8	381.4	216.7	56.8	215.0	98.4
200	771.0	37.8	4.9	770.4	424.8	55.1	423.1	99.2
Upper Salmon River								
10	38.3	8.6	22.5	39.2	24.6	62.8	23.1	88.2
20	76.4	12.3	16.1	79.1	41.8	52.8	39.9	91.1
50	129.1	20.1	15.6	197.9	91.8	46.4	89.6	95.3
100	384.2	28.3	7.4	385.4	174.6	45.3	172.3	97.4
200	770.3	39.1	5.1	792.3	335.1	42.3	332.8	98.6
500	1,921.9	61.5	3.2	1,917.8	790.8	41.2	788.4	99.4
1,000	3,850.4	90.2	2.3	3,872.2	1,661.3	42.9	1,658.8	99.7

[a] (SE of DTPH/SE of ETPH)^2.

Because σ^2_{DTPH} represents the within-day component of total variance of ETPH (σ^2_{ETPH}) for all days, then an alternative way of estimating σ^2_{DTPH} would be to subtract the day-to-day variance of ATPH (σ^2_{ATPH} = 72.25) in Figure 6A from the total variance of ETPH (σ^2_{ETPH} = 817.96) in Figure 6B. This gives a value of 745.71, as compared with the value of 750.76 obtained with the direct simulation of DTPH. The within-day variance (σ^2 = 745.71) represents the dominant component of to total variance (σ^2 = 817.96). The variance of ATPH (σ^2 = 72.25), the among-day component, represents a minor part of the total variance of ETPH, 817.96, whereas the within-day variance of DTPH represents (745.29/817.96)100 = 91.1% of the total variance of ETPH.

Other single-tributary routes.—The sampling distribution of our new estimator was also nonnormal for higher fishing effort on the Sterling River and Upper Salmon River routes, although it became progressively closer to normal with increasing effort. Table 1 shows the values already illustrated for the Sterling River route with 10 parties and also for the other party levels of the Sterling River route simulations and for all levels for the Upper Salmon River route simulations. Note that columns (1) and (2) are the values of the mean and standard deviation of ATPH illustrated by the example in Figure 6A, and columns (4) and (5) are the mean and standard error for ETPH (Figure 6B). When we graphed the corresponding histograms for these other levels of effort for the Sterling River and the Upper Salmon River routes, the histograms of ETPH became progressively less skewed with corresponding increases in fishing effort. For example, the coefficient of skewness (g) decreased from $g \approx 0.5$ for 10 parties on the Sterling River route to $g \approx 0.1$ for 200 parties. For the Upper Salmon River route, the skewness decreased from $g \approx 0.8$ for 10 parties to $g \approx 0.0$ for 1,000 parties.

The within-day variance component of ETPH (square of values in column 7 in Table 1) increased in magnitude relative to the total variance of ETPH (square of values in column 5), accounting for 91.1% (column 8) of the total variance for 10 parties on the Sterling River route and increasing to 99.2% for 200 parties. The same relationship held for the Upper Salmon River route. This phenomenon was associated with an increase in the variance of ATPH, the between-day variance (square of values in column 2), which was proportional to the increase in fishing effort, whereas the within-day variance of DTPH increased at a much faster rate.

Within increasing fishing effort, the coefficient of variation (100 × SE/mean) of ETPH (Table 1, column 6) stabilized at about 55% for the Sterling River and 43% for the Upper Salmon River routes. This stabilization indicates that a single functional form, such as the gamma or lognormal, might be used to fit the sampling distribution for each trib-

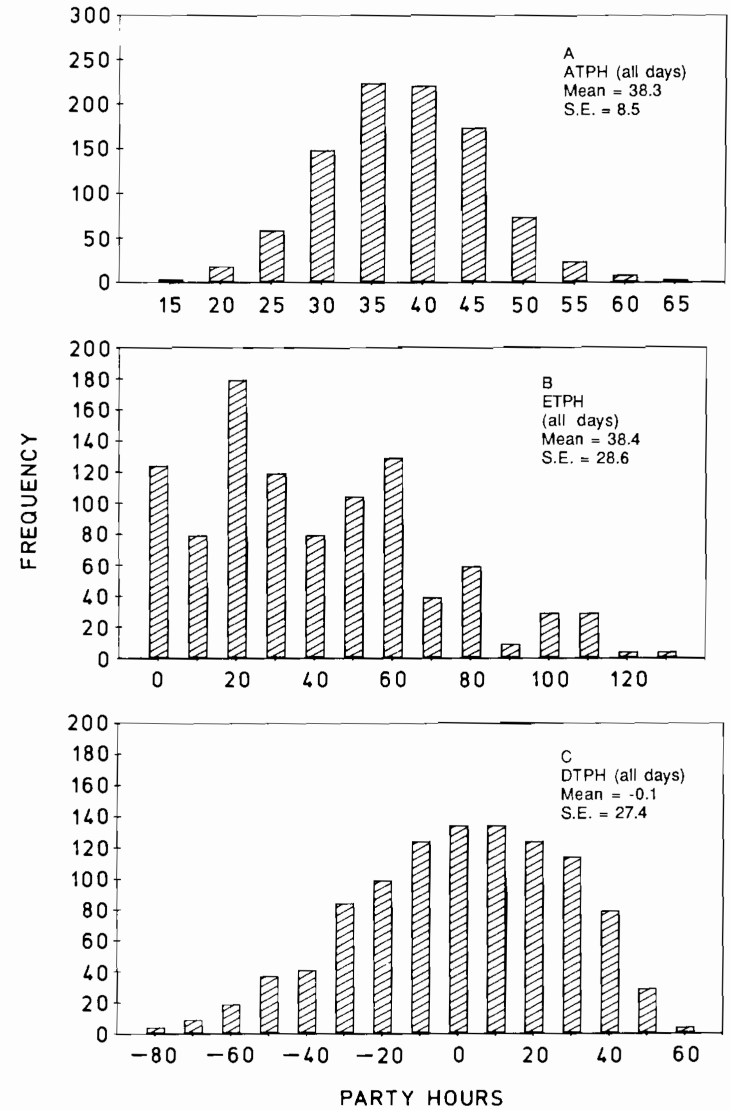

FIGURE 6.—Simulated behavior of the estimator of fishing effort of 10 parties for "multiple days" on the Sterling River tributary route. (**A**) Histogram of 1,000 stochastically produced values of actual total party hours (ATPH) was produced from 1,000 iterations of the angler module. (**B**) One thousand different estimations produced the histogram of total party hours (ETPH) that match each of the 1,000 iterations in (A). (**C**) The difference (DPTH) between the estimated and actual party hours shows the magnitude of the within-day variability.

utary route at all levels of effort. Either of these two parametric distribution functions is uniquely specified by the mean and the standard deviation of the sampling distribution. Specifically, the sampling distribution of ETPH was compared to a gamma distribution with the same mean and standard deviation for both the Sterling River (Figure 7) and Upper Salmon River routes. The potential benefits of fitting the gamma distribution are discussed later.

Behavior of the Estimator for Combined Tributary Routes

Simulations of weekend-day fishing effort.—The sampling distribution of the estimator continued to become more normal as routes were combined under weekend-day conditions. For illustration purposes, results for 10 parties on the Sterling River route and 100 parties on the Upper Salmon River route are presented in Figure 8, which is analogous

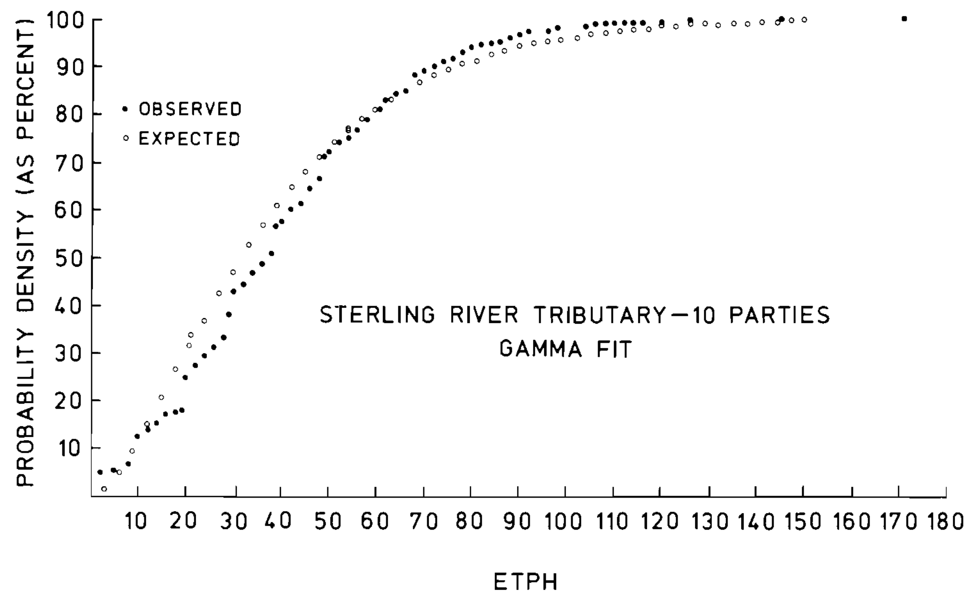

FIGURE 7.—The fit of the gamma distribution function with the observed values for estimated total party hours (ETPH) produced from simulations of the Sterling River tributary route with 10–200 parties. Expected values were generated from the gamma function. Observed values were produced from the simulations.

to Figure 6 for a single tributary route, but Figure 8 combines the Sterling route with the Upper Salmon route for eight weekend days within a month. This mirrors the sampling situation in which both weekend days are sampled each week. As in the single-tributary case, ETPH (Figure 8B) was an unbiased estimator of ATPH (Figure 8A) and had a broader range. Note that it had a more normal distribution than seen in the single-tributary case (Figure 6B). This was expected from the Central Limit Theorem. The within-day error estimation (DTPH) shown in Figure 8C had a mean of -1.1, not significantly different from 0.0 ($P > 0.05$). The distribution of DTPH more closely resembled that of ETPH (Figure 8B) as a result of the greater dominance of the within-day variability shown for a single day and a single tributary route in Table 1, column (8). Hence, under the conditions for which this new estimator was designed (i.e., multiple routes) its sampling distribution approaches that of a normal distribution.

Simulations of weekday fishing effort.—Simulation results for weekday sampling conditions mirrored those seen for weekend days. For illustrative purposes, the results for 10 parties on the Sterling River route and 100 parties on the Upper Salmon River route (combined data) are shown in Figure 9. The mean ATPH for this set of simulations was 423.7 (Figure 9A). The standard error of 13.4 was only slightly larger than for weekend-

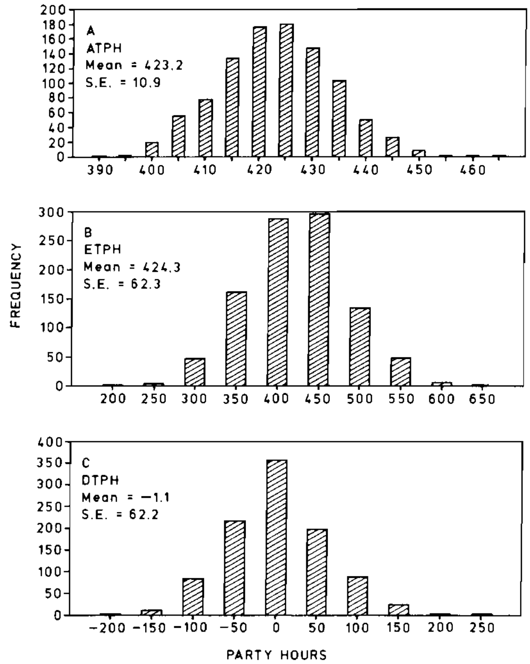

FIGURE 8.—Simulations of combined weekend-day fishing from the Sterling River (10 parties) and the Upper Salmon River (100 parties) tributary routes. These histograms illustrate weekend-day sampling regimes. One thousand iterations produced values for (**A**) actual total party hours, ATPH; (**B**) estimated total party hours, ETPH; and (**C**) the difference between ETPH and ATPH, defined as DTPH.

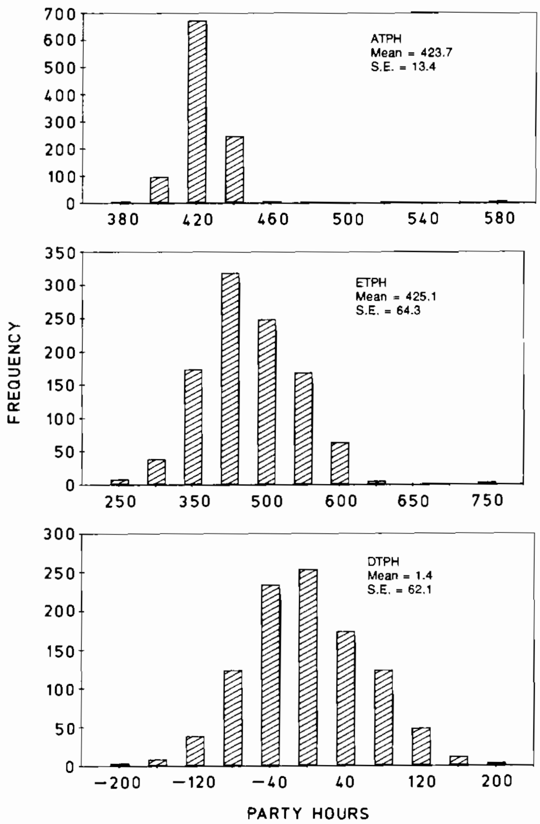

FIGURE 9.—Simulations for combined weekday fishing from the Sterling River (10 parties) and the Upper Salmon River (100 parties) tributary routes. These histograms illustrate weekday sampling regimes. One thousand iterations produced values for (**A**) actual total party hours, ATPH; (**B**) estimated total party hours, ETPH; and (**C**) the difference between ETPH and ATPH, defined as DTPH.

day simulations (Figure 8A). The mean ETPH of 425.1 (Figure 9B) was an unbiased estimate of mean ATPH with the same magnitude of standard error, 64.3 as for weekend days (Figure 8B). The mean DTPH (1.4) was not significantly different from 0.0 ($P > 0.05$). Again, as seen for the weekend-day analysis, the estimator became more normally distributed over combined routes than for the single-route case.

Confidence Interval Coverage

This new estimator's sampling distribution approached normality as more routes were combined in the final estimate of fishing effort. However, depending on whether one route or several were used, standard t-values based on the normal distribution can cause misspecification of the α values and confidence interval coverage. For these reasons, t-values for this estimator were determined empirically, based on sampling scenarios likely to be encountered.

Weekend-day simulations.—The near-normality of the sampling distribution of ETPH for the weekend-day simulations (Figure 8B) offered promise that confidence interval estimation of the conventional form

$$\text{Estimate} \pm t[\text{SE(Estimate)}]$$

could be used. To examine the validity of this procedure, we calculated the t-statistic for each of the two strata (weekday versus weekend day) of combined Sterling River and Upper Salmon River route simulations. For this t is the 8-d mean DTPH (Figure 8C) divided by the standard error calculated from the eight corresponding values of ETPH:

$$t = \frac{\frac{1}{8}\sum_{i=1}^{8} \text{DTPH}}{\text{SE(ETPH)}};$$

$$[\text{SE(ETPH)}]^2 = \frac{1}{64}\left[\sum_{i=1}^{8} (\text{ETPH})^2 - \frac{1}{8}\left(\sum_{i=1}^{8} \text{ETPH}\right)^2\right]$$

The sampling distribution of t is displayed as a cumulative distribution function in Figure 10, which shows the detail of the lower 50 t-values and upper 50 t-values. The actual t-values that were generated were compared with critical values from the Student t table, which is based on the assumption of a normally distributed estimator. For a Student t table, a value of ± 2.365 at the $P_\alpha = 0.05$ level yields 37 observations in the lower tail and 25 in the upper tail of our estimator's distribution, rather than the expected 25 in each tail. Hence, for accuracy, we generated the actual critical values from our simulated distributions. From these t-distributions (Figure 10), the actual values of t corresponding to the chosen P_α level can be read directly from the figure for use in building confidence intervals. For example, a $P_\alpha = 0.05$ with a two-tailed test would yield 25 observations in each tail out of a total of 1,000 observations. In this case, the lower tail t-value of -2.70 and the upper tail t-value of $+2.30$ (Figure 10) yield 25 observations in each tail.

Weekday simulations.—We used the same procedures to build the t frequency distribution for weekend days and for weekday simulations. Fig-

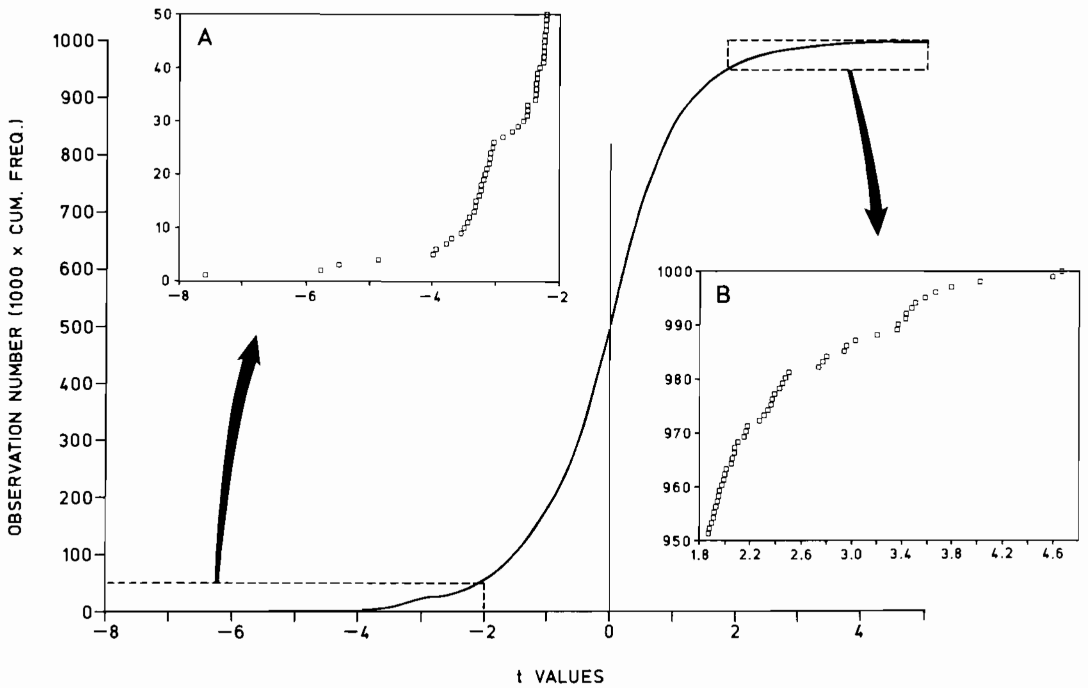

FIGURE 10.—The calculated cumulative distribution function of t-values (of the mean estimated total party hours) for weekend-day sampling. Inserts (A) and (B) show the details for the lower and upper 50 observations from this distribution.

ure 11 shows detail of the lower and upper 50 t-values. For the Student t table value of ± 2.365 (df = 7, P_α = 0.05), the lower tail contains 32 observations, and the upper tail contains 20 observations, rather than the expected 25 in each tail. The actual critical values that produce a P_α = 0.05 in a two-tailed test are -2.54 for the lower tail and $+2.25$ for the upper tail. These calculated t-values yield 25 observations in each tail.

Discussion

Application of this simulation model proved to be a useful tool to assess the statistical behavior of a new estimator of fishing effort. The estimator was shown to be unbiased (Robson and Jones 1989). However, the underlying form of the estimator's sampling distribution was unknown and was found through this numerical simulation to be non-normal. For simulation application purposes, this method gave an estimate of daily total fishing effort (ETPH) along a route on a given day. The sampling distribution in this circumstance was non-normal. However, when used in an actual angler survey, the mean ETPH for all sampling days within the time stratum, say weekend days of the month, is the statistic that is used to expand daily effort to monthly effort. We used simulations to show that normality increased with increasing ETPH, both within a route and, more importantly, when routes were pooled. This was encouraging because it suggests that the sampling distribution of the estimator will become normal over the larger fishery. The combined route case used here (i.e., the Sterling and Upper Salmon routes) was based on independent simulations of these two tributary routes. This creates somewhat of a false impression because the approach to normality in field studies will not be as fast. However, in the case of the New York Great Lakes angler survey, 10 tributary routes were combined to yield a lakewide estimate of stream fishing effort, and our current simulations may be very realistic.

The coefficients of variation from the simulations were fairly homogeneous among the tributary routes. This was a necessary condition for pooling tributary route effort estimates into a lakewide estimate of effort. When the coefficients are similar, it is likely that the sampling distributions for route estimations are the same. This estimator was built specifically to be used for large geographical areas broken into several survey routes. The

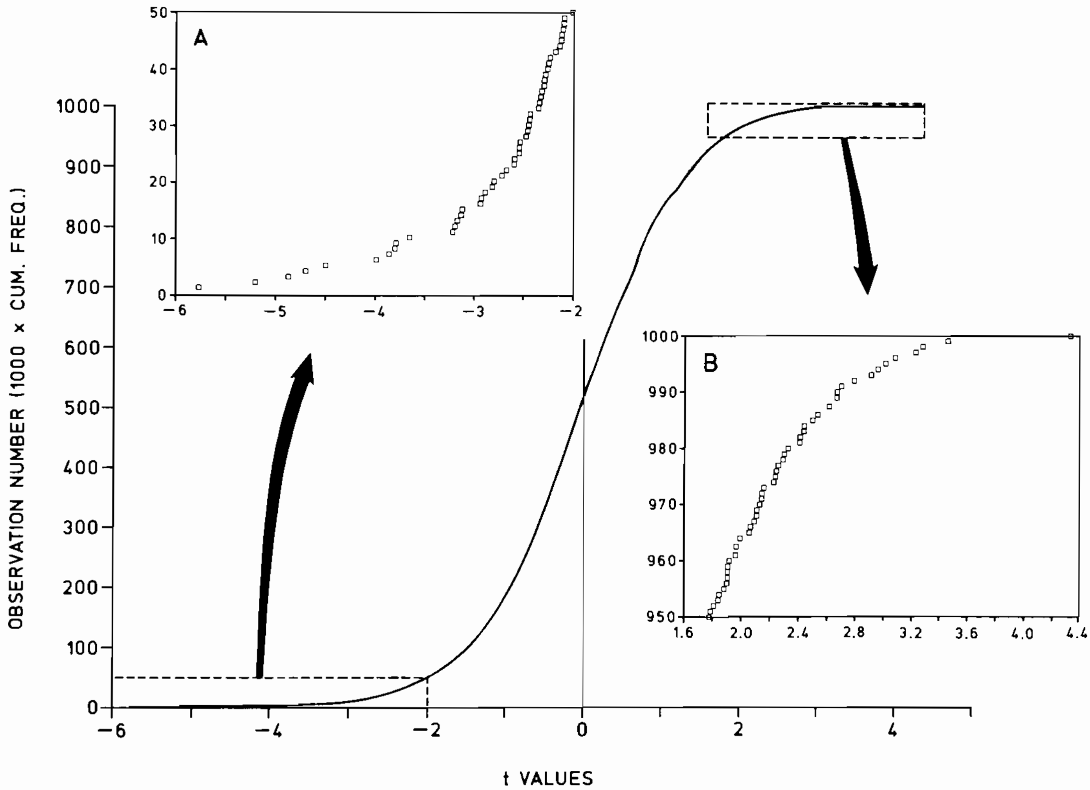

FIGURE 11.—The calculated cumulative distribution function of t-values (of the mean estimated total party hours) for weekday sampling. Inserts (A) and (B) show the details for the lower and upper 50 observations from this distribution.

estimator works best when pooled over routes from the same fishery (e.g., tributary fishing).

The simulations revealed valuable information concerning the magnitude of the variance components. It showed that, for the New York fishery, the within-day variance dominated the total variance, and the day-to-day variance component was of negligible magnitude. This condition would not necessarily hold for all fisheries and depends on the heterogeneity of fishing over the season.

The simulations also allowed calculation of DTPH and therefore insight into the distribution of t-values. Critical (t) values can be taken directly from the figures given in this paper. Additionally, the simulations have shown that use of the table t-values based on a normal distribution will result in misspecification of the P_α region in the calculation of confidence limits.

The simulation model also revealed that the sampling distribution of ETPH could be approximated with a gamma distribution. Hence, as an aid in planning similar future studies, the daily ETPH sampling distribution for a route might be assumed to have the shape of a gamma distribution with a CV on the order of 50% and a mean value equal to the actual daily mean ATPH. This information also suggests the possibility of using the gamma distribution for confidence interval estimation rather than using the empirically determined t-distribution. Once a known distribution function can be fit to the empirical sampling distribution, mathematic methods exist for the solution of the critical values for specific alpha levels of confidence. The critical intervals for the gamma distribution can be found by using the large-scale approximate confidence interval method described in Mendenhall et al. (1981).

Acknowledgments

This work was supported by U.S. Fish and Wildlife Service, Cooperative Agreement 14-16-0009-1528-RW04 with Cornell University, New York; by the New York Cooperative Fish and Wildlife Research Unit, Cornell University; and by the New York State Department of Environmental Conservation, contract C001257. We thank

Marcel Hawiger for his excellent programming assistance.

References

Diaconis, P., and B. Efron. 1983. Computer-intensive methods in statistics. Scientific American 248(5): 116–130.

Efron, B. 1979a. Bootstrap methods: another look at the jackknife. Annals of Statistics 7:1–26.

Efron, B. 1979b. Computers and the theory of statistics: thinking the unthinkable. SIAM (Society for Industrial and Applied Mathematics) Review 21: 460–480.

Law, A. M., and W. D. Kelton. 1982. Simulation modeling and analysis. McGraw-Hill, New York.

Mendenhall, W., R. L. Scheaffer, and D. D. Wackerly. 1981. Mathematical statistics with applications. Duxbury Press, Boston.

Morgan, B. J. T. 1984. Elements of simulation. Chapman and Hall, New York.

Robson, D. S., and C. M. Jones. 1989. The theoretical basis of an access site angler survey design. Biometrics 45:83–98.

Received April 7, 1989
Accepted September 18, 1989